

Shadow Removal via Shadow Image Decomposition

Hieu Le
 Stony Brook University
 New York, 11794, USA
 hle@cs.stonybrook.edu

Dimitris Samaras
 Stony Brook University
 New York, 11794, USA
 samaras@cs.stonybrook.edu

Abstract

We propose a novel deep learning method for shadow removal. Inspired by physical models of shadow formation, we use a linear illumination transformation to model the shadow effects in the image that allows the shadow image to be expressed as a combination of the shadow-free image, the shadow parameters, and a matte layer. We use two deep networks, namely SP-Net and M-Net, to predict the shadow parameters and the shadow matte respectively. This system allows us to remove the shadow effects on the images. We train and test our framework on the most challenging shadow removal dataset (ISTD). Compared to the state-of-the-art method, our model achieves a 40% error reduction in terms of root mean square error (RMSE) for the shadow area, reducing RMSE from 13.3 to 7.9. Moreover, we create an augmented ISTD dataset based on an image decomposition system by modifying the shadow parameters to generate new synthetic shadow images. Training our model on this new augmented ISTD dataset further lowers the RMSE on the shadow area to 7.4.

1. Introduction

Shadows are cast whenever a light source is blocked by an object. Shadows often confound computer vision algorithms such as segmentation, tracking, or recognition. The appearance of shadow edges is hard to distinguish from edges due to material changes [27]. Dark albedo material regions can be easily misclassified as shadows [18]. Thus many methods have been proposed to identify and remove shadows from images.

Early shadow removal work was based on physical shadow models [1]. A common approach is to formulate the shadow removal problem using an image formation model, in which the image is expressed in terms of material properties and a light source-occluder system that casts shadows. Hence, a shadow-free image can be obtained by estimating the parameters of the source-occluder system and then reversing the shadow effects on the image [10, 14, 13, 28].

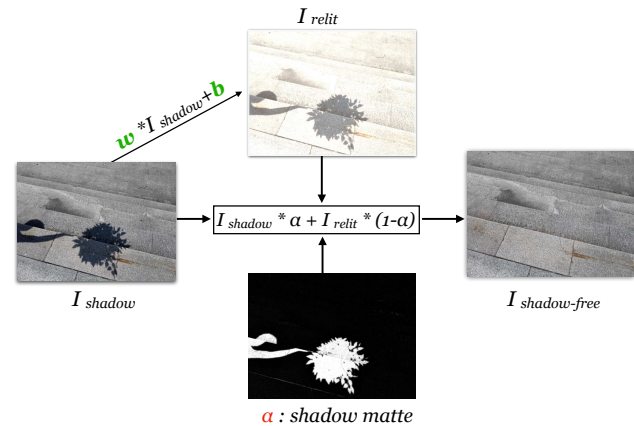


Figure 1: **Shadow Removal via Shadow Image Decomposition.** A shadow-free image $I_{\text{shadow-free}}$ can be expressed in terms of a shadow image I_{shadow} , a relit image I_{relit} and a shadow matte α . The relit image is a linear transformation of the shadow image. The two unknown factors of this system are the shadow parameters (w, b) and the shadow matte layer α . We use two deep networks to estimate these two unknown factors.

These methods relight the shadows in a physically plausible manner. However, estimating the correct solution for such illumination models is non-trivial and requires considerable processing time or user assistance[39, 3].

On the other hand, recently published large-scale datasets [25, 34, 32] allow the use of deep learning methods for shadow removal. In these cases, a network is trained in an end-to-end fashion to map the input shadow image to a shadow-free image. The success of these approaches shows that deep networks can effectively learn transformations that relight shadowed pixels. However, the actual physical properties of shadows are ignored, and there is no guarantee that the networks would learn physically plausible transformations. Moreover, there are still well known

issues with images generated by deep networks: results tend to be blurry [15, 40] and/or contain artifacts [23]. How to improve the quality of generated images is an active research topic [16, 35].

In this work, we propose a novel method for shadow removal that takes advantage of both shadow illumination modelling and deep learning. Following early shadow removal works, we propose to use a simplified physical illumination model to define the mapping between shadow pixels and their shadow-free counterparts.

Our proposed illumination model is a linear transformation consisting of a scaling factor and an additive constant - per color channel - for the whole umbra area of the shadow. These scaling factors and additive constants are the parameters of the model, see Fig. 1. The illumination model plays a key role in our method: with correct parameter estimates, we can use the model to remove shadows from images. We propose to train a deep network (SP-Net) to automatically estimate the parameters of the shadow model. Through training, SP-Net learns a mapping function from input shadow images to illumination model parameters.

Furthermore, we use a shadow matting technique [3, 13, 39] to handle the penumbra area of the shadows. We incorporate our illumination model into an image decomposition formulation [24, 3], where the shadow-free image is expressed as a combination of the shadow image, the parameters of the shadow model, and a shadow density matte. This image decomposition formulation allows us to reconstruct the shadow-free image, as illustrated in Fig. 1. The shadow parameters (w, b) represent the transformation from the shadowed pixels to the illuminated pixels. The shadow matte represents the per-pixel linear combination of the relit image and the shadow image, which results to the shadow-free image. Previous work often requires user assistance [12] or solving an optimization system [20] to obtain the shadow mattes. In contrast, we propose to train a second network (M-Net) to accurately predict shadow mattes in a fully automated manner.

We train and test our proposed SP-Net and M-Net on the ISTD dataset [34], which is the largest and most challenging available dataset for shadow removal. SP-Net alone (no matting) outperforms the state-of-the-art [12] in shadow removal by 29% in terms of RMSE on shadow areas, from 13.3 to 9.5 RMSE. Our full system with both SP-Net and M-Net further improves the overall results by another 17%, which yields a RMSE of 7.9.

Our proposed method can realistically modify the shadow effects in the images. First we estimate the shadow parameters and shadow matte from an image. We then add the shadows back into the shadow-free image with a set of modified shadow parameters. As we change the parameters, the shadow effects change accordingly. In this manner, we can synthesize additional shadow images that serve as aug-

mented training data. Training our system on ISTD plus our newly synthesized images further lowers the RMSE on the shadow areas by 6%, compared to our model trained on the original ISTD dataset.

The main contributions of this work are:

- We propose a new deep learning approach for shadow removal, grounded by a simplified physical illumination model and an image decomposition formulation.
- We propose a method for shadow image augmentation based on our simplified physical illumination model and the image decomposition formulation.
- Our proposed method achieves state-of-the-art shadow removal results on the ISTD dataset.

The pre-trained model, shadow removal results, and more details can be found at: www3.cs.stonybrook.edu/~cvl/projects/SID/index.html

2. Related Works

Shadow Illumination Models: Early research on shadow removal is motivated by physical modelling of illumination and color [10, 9, 11, 6]. Barrow & Tenenbaum [1] define an intrinsic image algorithm that separates images into the intrinsic components of reflectance and shading. Guo *et al.* [13] simplify this model to represent the relationship between the shadow pixels and shadow-free pixels via a linear system. They estimate the unknown factors via pairing shadow and shadow-free regions. Similarly, Shor & Lischinski [28] propose an illumination model for shadows in which there is an affine relationship between the lit and shadow intensities at a pixel, including 4 unknown parameters. They define two strips of pixels: one in the shadowed area and one in the lit area to estimate their parameters. Finlayson *et al.* [8] create an illuminant-invariant image for shadow detection and removal. Their work is based on an insight that the shadowed pixels differ from their lit pixels by a scaling factor. Vicente *et al.* [31, 33] propose a method for shadow removal where they suggest that the color of the lit region can be transferred to the shadowed region via histogram equalization.

Shadow Matting: Matting, introduced by Porter & Duff [24], is an effective tool to handle soft shadows. However, it is non-trivial to compute the shadow matte from a single image. Chuang *et al.* [3] use image matting for shadow editing to transfer the shadows between different scenes. They compute the shadow matte from a sequence of frames in a video captured from a static camera. Guo *et al.* [13] and Zhang *et al.* [39] both use a shadow matte for their shadow removal frameworks, where they estimate the shadow matte via the closed-form solution of Levin *et al.* [20].

Deep-Learning Based Shadow Removal: Recently published large-scale datasets [32, 34, 25] enable training deep-learning networks for shadow removal. The Deshadow-Net of Qu *et al.* [25] is trained to remove shadows in an end-to-end manner. Their network extracts multi-context features across different layers of a deep network to predict a shadow matte. This shadow matte is different from ours as it contains both the density and color offset of the shadows. The ST-CGAN proposed by Wang *et al.* [34] for both shadow detection and removal is a conditional GAN-based framework [15] for shadow detection and removal. Their framework is trained to predict the shadow mask and shadow-free image in a unified manner, they use GAN losses to improve performance.

Inspired by early work, our framework outputs the shadow-free image based on a physically inspired shadow illumination model and a shadow matte. We, however, estimate the parameters of our model and the shadow matte via two deep networks in a fully automated manner.

3. Shadow and Image Decomposition Model

3.1. Shadow Illumination Model

Let us begin by describing our shadow illumination model. We aim to find a mapping function T to transform a shadow pixel I_x^{shadow} to its non-shadow counterpart: $I_x^{\text{shadow-free}} = T(I_x^{\text{shadow}}, w)$ where w are the parameters of the model. The form of T has been studied in depth in previous work as discussed in Sec. 2.

In this paper, similar to the model of Shor & Lischinski [28], we use a linear function to model the relationship between the lit and shadowed pixels. The intensity of a lit pixel is formulated as:

$$I_x^{\text{shadow-free}}(\lambda) = L_x^d(\lambda)R_x(\lambda) + L_x^a(\lambda)R_x(\lambda) \quad (1)$$

where $I_x^{\text{shadow-free}}(\lambda)$ is the intensity reflected from point x in the scene at wavelength λ , L and R are the illumination and reflectance respectively, L^d is the direct illumination and L^a is the ambient illumination.

To cast a shadow on point x , an occluder blocks the direct illumination and a portion of the ambient illumination that would otherwise arrive at x . The shadowed intensity at x is:

$$I_x^{\text{shadow}}(\lambda) = a_x(\lambda)L_x^a(\lambda)R_x(\lambda) \quad (2)$$

where $a_x(\lambda)$ is the attenuation factor indicating the remaining fraction of the ambient illumination that arrives at point x at wavelength λ . Note that Shor & Lischinski further assume that $a_x(\lambda)$ is the same for all wavelengths λ to simplify their model. This assumption implies that the environment light has the same color from all directions.

From Eq. 1 and 2, we can express the shadow-free pixel as a linear function of the shadowed pixel:

$$I_x^{\text{shadow-free}}(\lambda) = L_x^d(\lambda)R_x(\lambda) + a_x(\lambda)^{-1}I_x^{\text{shadow}}(\lambda) \quad (3)$$

We assume that this linear relation is preserved throughout the color acquisition process of the camera [7]. Therefore, we can express the color intensity of the lit pixel x as a linear function of its shadowed value:

$$I_x^{\text{shadow-free}}(k) = w_k \times I_x^{\text{shadow}}(k) + b_k \quad (4)$$

where $I_x(k)$ represents the value of the pixel x on the image I in color channel k ($k \in R, G, B$ color channel), b_k is the response of the camera to direct illumination, and w_k is responsible for the attenuation factor of the ambient illumination at this pixel in this color channel. We model each color channel independently to account for possibly different spectral characteristics of the material in shadow as well as the sensor.

We further assume that the two vectors $w = [w_R, w_G, w_B]$ and $b = [b_R, b_G, b_B]$ are constant across all pixels x in the umbra area of the shadow. Under this assumption, we can easily estimate the values of w and b given the shadow and shadow-free image using linear regression. We refer to (w, b) as the *shadow parameters* in the rest of the paper.

In Sec. 4, we show that we can train a deep-network to estimate these vectors from a single image.

3.2. Shadow Image Decomposition System

We plug our proposed shadow illumination model into the following well-known image decomposition system [3, 24, 30, 36]. The system models the shadow-free image using the shadow image, the shadow parameter, and the shadow matte. The shadow-free image can be expressed as:

$$I^{\text{shadow-free}} = I^{\text{shadow}} \cdot \alpha + I^{\text{relit}} \cdot (1 - \alpha) \quad (5)$$

where I^{shadow} and $I^{\text{shadow-free}}$ are the shadow and shadow-free image respectively, α is the matting layer, and I^{relit} is the relit image. We define α and I^{relit} below.

Each pixel i of the relit image I^{relit} is computed by:

$$I_i^{\text{relit}} = w \cdot I_i^{\text{shadow}} + b \quad (6)$$

which is the shadow image transformed by the illumination model of Eq. 4. This transformation maps the shadowed pixels to their shadow-free values.

The matting layer α represents the per-pixel coefficients of the linear combination of the relit image and the input shadow image that results into the shadow-free image. Ideally, the value of α should be 1 at the non-shadow area and 0 at the umbra of the shadow area. For the pixels in the penumbra of the shadow, the value of α gradually changes near the shadow boundary.

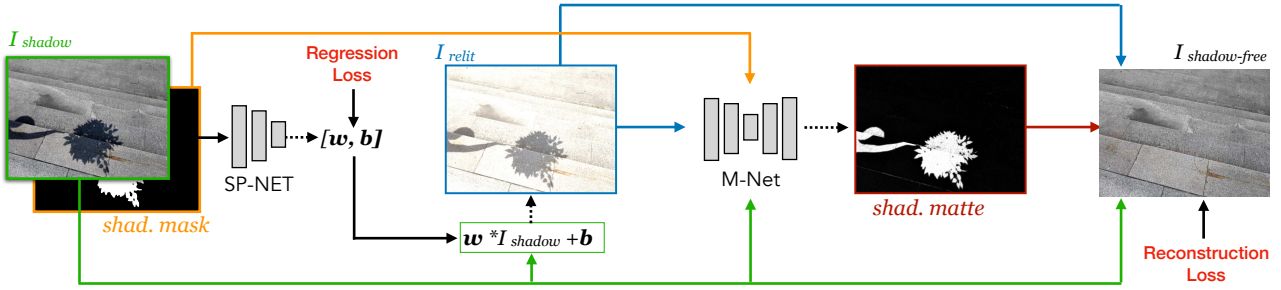


Figure 2: **Shadow Removal Framework.** The shadow parameter estimator network SP-Net takes as input the shadow image and the shadow mask to predict the shadow parameters (w, b) . The relit image I^{relit} is then computed via Eq. 6 using the estimated parameters from SP-Net. The relit image, together with the input shadow image and the shadow mask are then input into the shadow matte prediction network M-Net to get the shadow matte layer α . The system outputs the shadow-free image via Eq. 5, using the shadow image, the relit image, and the shadow matte. SP-Net learns to predict the shadow parameters (w, b) , denoted as the regression loss. M-Net learns to minimize the L_1 distance between the output of the system and the shadow-free image (reconstruction loss).

The value of α at pixel i based on the shadow image, shadow-free image, and relit image, follows from Eq. 5 :

$$\alpha_i = \frac{I_i^{\text{shadow-free}} - I_i^{\text{relit}}}{I_i^{\text{shadow}} - I_i^{\text{relit}}} \quad (7)$$

We use the image decomposition of Eq. 5 for our shadow removal framework. The unknown factors are the shadow parameters (w, b) and the shadow matte α . We present our method that uses two deep networks, SP-Net and M-Net, to predict these two factors in the following section. In Sec.5.3, we propose a simple method to modify the shadows for an image in order to augment the training data.

4. Shadow Removal Framework

Fig. 2 summarizes our framework. The shadow parameter estimator network SP-Net takes as input the shadow image and the shadow mask to predict the shadow parameters (w, b) . The relit image I^{relit} is then computed via Eq. 6 with the estimated parameters from SP-Net. The relit image, together with the input shadow image and the shadow mask is then input into the shadow matte prediction network M-Net to get the shadow matte α . The system outputs the shadow-free image via Eq. 5.

4.1. Shadow Parameter Estimator Network

In order to recover the illuminated intensity at the shadowed pixel, we need to estimate the parameters of the linear model in Eq. 4. Previous work has proposed different methods to estimate the parameters of a shadow illumination model [28, 12, 13, 11, 8, 6]. In this paper, we train SP-Net, a deep network model, to directly predict the shadow parameters from the input shadow image.

To train SP-Net, we first generate training data. Given a training pair of a shadow image and a shadow-free image, we estimate the parameters of our linear illumination model using a least squares method [4]. For each shadow image, we first erode the shadow mask by 5 pixels in order to define a region that does not contain the partially shadowed (penumbra) pixels. Mapping these shadow pixel values to the corresponding values in the shadow-free image, gives us a linear regression system, from which we calculate w and b . We compute parameters for each of the three RGB color channels and then combine the learned coefficients to form a 6-element vector. This vector is used as the targeted output to train SP-Net. The input for SP-Net is the input shadow image and the associated shadow mask. We train SP-Net to minimize the L_1 distance between the output of the network and these computed shadow parameters.

We develop SP-Net by customizing a ResNeXt [37] model that is pre-trained on ImageNet [5]. Notice that while we use the ground truth shadow mask for training, during testing we estimate shadow masks using the shadow detection network proposed by Zhu *et al.*[41].

4.2. Shadow Matte Prediction Network

Our linear illumination model (Eq. 4) can relight the pixels in the umbra area (fully shadowed). The shadowed pixels in the penumbra (partially shadowed) region are more challenging as the illumination changes gradually across the shadow boundary [14]. A binary shadow mask cannot model this gradual change. Thus, using a binary mask within the decomposition model in Eq. 5 will generate an image with visible boundary artifacts. A solution for this is shadow matting where the soft shadow effects are expressed via the values of a blending layer.

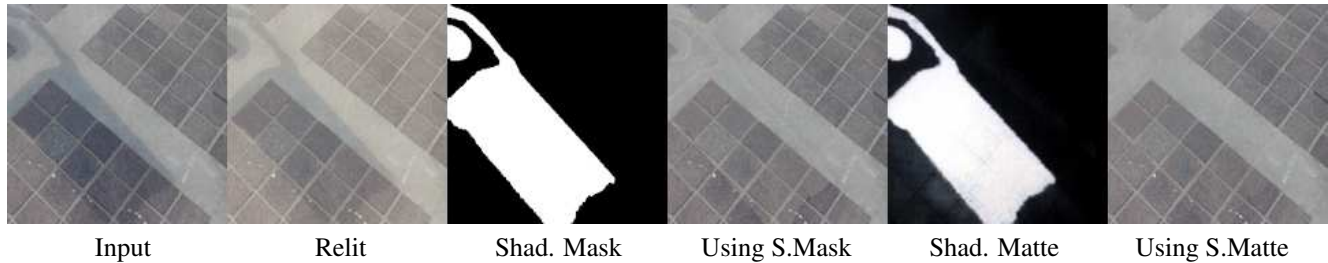


Figure 3: **A comparison of the ground truth shadow mask and our shadow matte.** From the left to right: The input image, the relit image computed from the parameters estimated via SP-Net, the ground truth shadow mask, the final results when we use the shadow mask, the shadow matte computed using our M-Net, and the final shadow-free image when we use the shadow matte to combine the input and relit image. The matting layer handles the soft shadow and does not generate visible boundaries in the final result. (*Please view in magnification on a digital device to see the difference more clearly.*)

In this paper, we train a deep network, M-Net, to predict this matting layer. In order to train M-Net, we use Eq. 5 to compute the output of our framework where the shadow matte is the output of M-Net. Then the loss function that drives the training of M-Net is the L_1 distance between output image and ground truth training shadow-free image, marked as “reconstruction loss” in Fig. 2. This is equivalent to computing the actual value of the shadow matte via Eq. 7 and then training M-Net to directly output this value.

Fig. 3 illustrates the effectiveness of our shadow matting technique. We show in the figure two shadow removal results which are computed using a ground-truth shadow mask and a shadow matte respectively. This shadow matte is computed by our model. One can see that using the binary shadow mask to form the shadow-free image creates visible boundary artifacts as it ignores the penumbra. The shadow matte from our model captures well the soft shadow and generates an image without shadow boundary artifacts.

We design M-Net based on U-Net [26]. The M-Net inputs are the shadow image, the relit image, and the shadow mask. We use the shadow mask as input to M-Net since the matting layer can be considered as a relaxed shadow mask where each value represents the strength of the shadow effect at the location rather than just the shadow presence.

5. Experiments

5.1. Dataset and Evaluation Metric

We train and evaluate on the ISTD dataset [34]. ISTD consists of image triplets: shadow image, shadow mask, and shadow-free image, captured from different scenes. The training split has 1870 image triplets from 135 scenes, whereas the testing split has 540 triplets from 45 scenes.

We notice that the testing set of the ISTD dataset needs to be adjusted since the shadow images and the shadow-free images have inconsistent colors. This is a well known issue mentioned in the original paper [34]. The reason is that the shadow and shadow-free image pairs were captured



Figure 4: **An example of our color correction method.** From left to right: input shadow image, provided shadow-free ground truth image (GT) from ISTD dataset, and the GT image corrected by our method. Comparing to the input shadow image on the non-shadow area only, the root-mean-square distance of the original GT is 12.9. This value on our corrected GT becomes 2.9.

at different times of the day which resulted in slightly different environment lights for each image. For example, Fig. 4 shows a shadow and shadow-free image pair. The root-mean-square difference between these two images in the non-shadow area is 12.9. This color inconsistency appears frequently in the testing set of the ISTD dataset. On the whole testing set, the root-mean-square distance between the shadow images and shadow-free images in the non-shadow area is 6.83, as computed by Wang *et al.*[34].

In order to mitigate this color inconsistency, we use linear regression to transform the pixel values in the non-shadow area of each shadow-free image to map into their counterpart values in the shadow image. We use a linear regression for each color-channel, similar to our method for relighting the shadow pixels in Sec. 4.1. This simple transformation transfers the color tone and brightness of the shadow image to its shadow-free counterpart. The third column of Fig. 4 illustrates the effect of our color-correction method. Our proposed method reduces the root-mean-square distance between the shadow-free image and the shadow image from 12.9 to 2.9. The error reduction for the whole testing set of ISTD goes from 6.83 to 2.6.

5.2. Shadow Removal Evaluation

We evaluate our method on the adjusted testing set of the ISTD dataset. For metric evaluation we follow [34] and compute the RMSE in the LAB color space on the shadow area, non-shadow area, and the whole image, where all shadow removal results are re-sized into 256×256 to compare with the ground truth images at this size. Note that in contrast to other methods that only output shadow free images at that resolution, our shadow removal system works for input images of any size. Since our method requires shadow masks, we use the model proposed by Zhu *et al.*[41] pre-trained on the SBU dataset [32] for detecting shadows. We take the model provided by the author and fine-tune it on the ISTD dataset for 3000 epochs. This model achieves 2.2 Balance Error Rate on the ISTD testing set. To remove the shadow effect in the image, we first use SP-Net to compute the shadow parameters (w, b) using the input image and the shadow mask computed from the shadow detection network. We use (w, b) to compute a re-lit image which is input to M-Net, together with the input image and the shadow mask to output a matte layer. We obtain the final shadow removal result via Eq. 5. In Table 1, we compare the performance of our method with the recent shadow removal methods of Guo *et al.*[13], Yang *et al.*[38], Gong *et al.*[12], and Wang *et al.*[34]. All numbers are computed on the adjusted testing images so that they are directly comparable. The first row shows the numbers for the input shadow images, i.e. no shadow removal performed.

We first evaluate our shadow removal performance using only SP-Net, i.e. we use the binary shadow mask computed by the shadow detector to form the shadow-free image from the shadow image and the relit image. The binary shadow mask is obtained by simply thresholding the output of the shadow detector with a threshold of 0.95. As shown in column “SP-Net” (third from the right) in Fig. 8, SP-Net correctly estimates the shadow parameters to relight the shadow area. Even with visible shadow boundaries, SP-Net alone outperforms the previous state-of-the-art, reducing the RMSE on the shadow area by 29%, from 13.3 to 9.5.

We then evaluate the shadow removal results using both SP-Net and M-Net, denoted as “SP+M-Net” in Tab. 1 and Fig. 8. As shown in Fig. 8, the results of M-Net do not contain boundary artifacts. In the third row of Fig. 8, SP-Net overly relights the shadow area but the shadow matte computed from M-Net effectively corrects these errors. This is because M-Net is trained to blend the relit and shadow images to create the shadow-free image. Therefore, M-Net learns to output a smaller weight for a pixel that is overly lit by SP-Net. Using the matte layer of M-Net further reduces the RMSE on the shadow area by 17%, from 9.5 to 7.9.

Overall, our method generates better results than other methods. Our method does a better job at estimating the

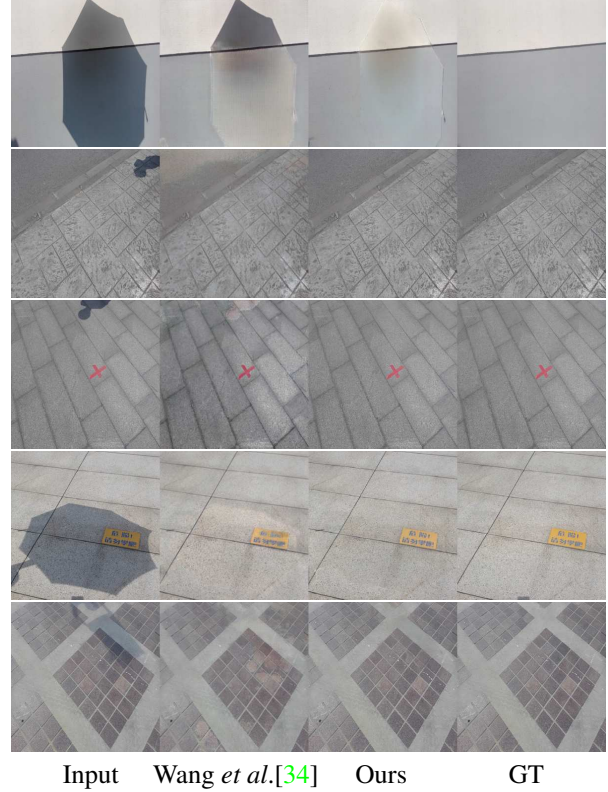


Figure 5: **Comparison of shadow removal between our method and ST-CGAN [34].** ST-CGAN tends to produce blurry images, random artifacts, and incorrect colors of the lit pixels while our method handles all cases well.

overall illumination changes compared to the model of Gong *et al.*, which tends to overly relight shadow pixels, as shown in Fig. 8. Our method does not show color inconsistencies within the relit area contrary to all other methods. Fig. 5 qualitatively compares our method and ST-CGAN, which illustrates common issues present in images generated by deep networks [15, 40]. ST-CGAN generally generates blurry images and introduces random artifacts. Our method, albeit not perfect, handles all cases well.

Our method fails to recover the shadow-free pixels properly as shown in Fig. 6. The first row, shows how our method overly relights the shadowed area while in the second row, the color of the lit area is incorrect.

Finally, we trained and evaluated two alternative designs that do not require shadow masks as input: (1) The first is an end-to-end shadow-removal system where we jointly train a shadow detector together with our proposed SP-Net and M-Net. This framework is harder to train due to the increase in the number of network parameters. (2) The second is a version of our framework that does not input the shadow masks into both SP-Net and M-Net. Hence, SP-Net and M-Net need to learn to localize the shadow areas implicitly.

Table 1: **Shadow removal results of our networks compared to state-of-the-art shadow removal methods on the adjusted ground truth.** (*) The method of Gong *et al.* [12] is an interactive method that defines the shadow/non-shadow regions via user inputs, thus generates minimal error on the non-shadow area. The metric is RMSE (the lower, the better). Best results are in bold.

Methods	Shadow	Non-Shadow	All
Input Image	40.2	2.6	8.5
Yang <i>et al.</i> [38]	24.7	14.4	16.0
Guo <i>et al.</i> [13]	22.0	3.1	6.1
Wang <i>et al.</i> [34]	13.4	7.7	8.7
Gong <i>et al.</i> [12]	13.3	2.6*	4.2
SP-Net (Ours)	9.5	3.2	4.1
SP+M-Net (Ours)	7.9	3.1	3.9
Our Method with Alternative Settings			
With a Shad. Detector	8.4	5.0	5.5
No Input Shadow Mask	8.3	4.9	5.4

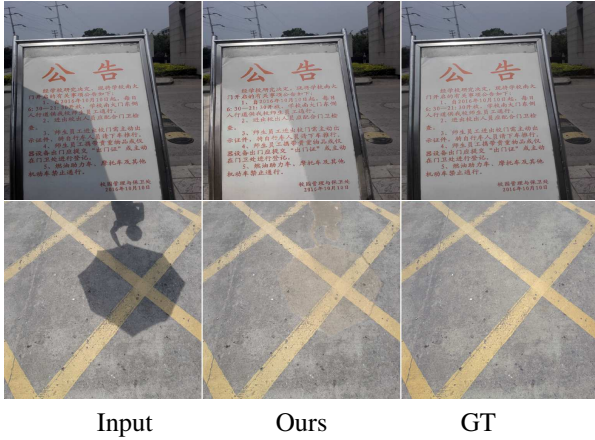


Figure 6: **Failure cases of our method.** In the first row, our method overly lights up the shadow area. In the second row, our method generates incorrect colors.

As can be seen in the two bottom rows of Tab. 1, both designs achieved slightly worse shadow removal results than our main setting.

5.3. Dataset Augmentation via Shadow Editing

Many deep learning work focus on learning from more easily obtainable, weakly-supervised, or synthetic data [2, 19, 21, 22, 29, 18, 17]. In this section, we show that we can modify shadow effects using our proposed illumination model to generate additional training data.

Given a shadow matte α , a shadow-free image, and pa-



Syns. Image Real Image Syns. Image
 $w_{\text{syn}} = w \times 0.8$ $w_{\text{syn}} = w \times 1.7$

Figure 7: **Shadow editing via our decomposition model.** We use Eq. 8 to generate synthetic shadow images. As we change the shadow parameters, the shadow effects change accordingly. We show two example images from the ISTD training set where in the middle column are the original images and in the first and last column are synthetic.

Table 2: **Shadow removal results of our networks train on the augmented ISTD dataset.** The metric is RMSE (the lower, the better). Training our framework on the augmented ISTD dataset drops the RMSE on the shadow area from 7.9 to 7.4.

Methods	Train. Set	Shad.	Non-Shad.	All
SP-Net	Aug. ISTD	9.0	3.2	4.1
SP+M-Net	Aug. ISTD	7.4	3.1	3.8

rameters (w, b) , we can form a shadow image by:

$$I^{\text{shadow}} = I^{\text{shadow-free}} \cdot \alpha + I^{\text{darkened}} \cdot (1 - \alpha) \quad (8)$$

where I^{darkened} has undergone the shadow effect associated to the set of shadow parameters (w, b) . Each pixel i of I^{darkened} is computed by:

$$I_i^{\text{darkened}} = (I_i^{\text{shadow-free}} - b) \cdot w^{-1} \quad (9)$$

For each training image, we first compute the shadow parameters and the matte layer via Eqs. 4 and 7. Then, we generate a new synthetic shadow image via Eq. 8 with a scaling factor $w_{\text{syn}} = w \times k$. As seen in Fig. 7, a lower w leads to an image with a lighter shadow area while a higher w increases the shadow effects instead. Using this method, we augment the ISTD training set by simply choosing $k = [0.8, 0.9, 1.1, 1.2]$ to generate a new set of 5320 images, which is four times bigger than the original training set. We augment the original ISTD dataset with this dataset. Training our model on this new augmented ISTD dataset improves our results, as the RMSE drops by 6%, from 7.9 to 7.4, as reported in Tab. 2.

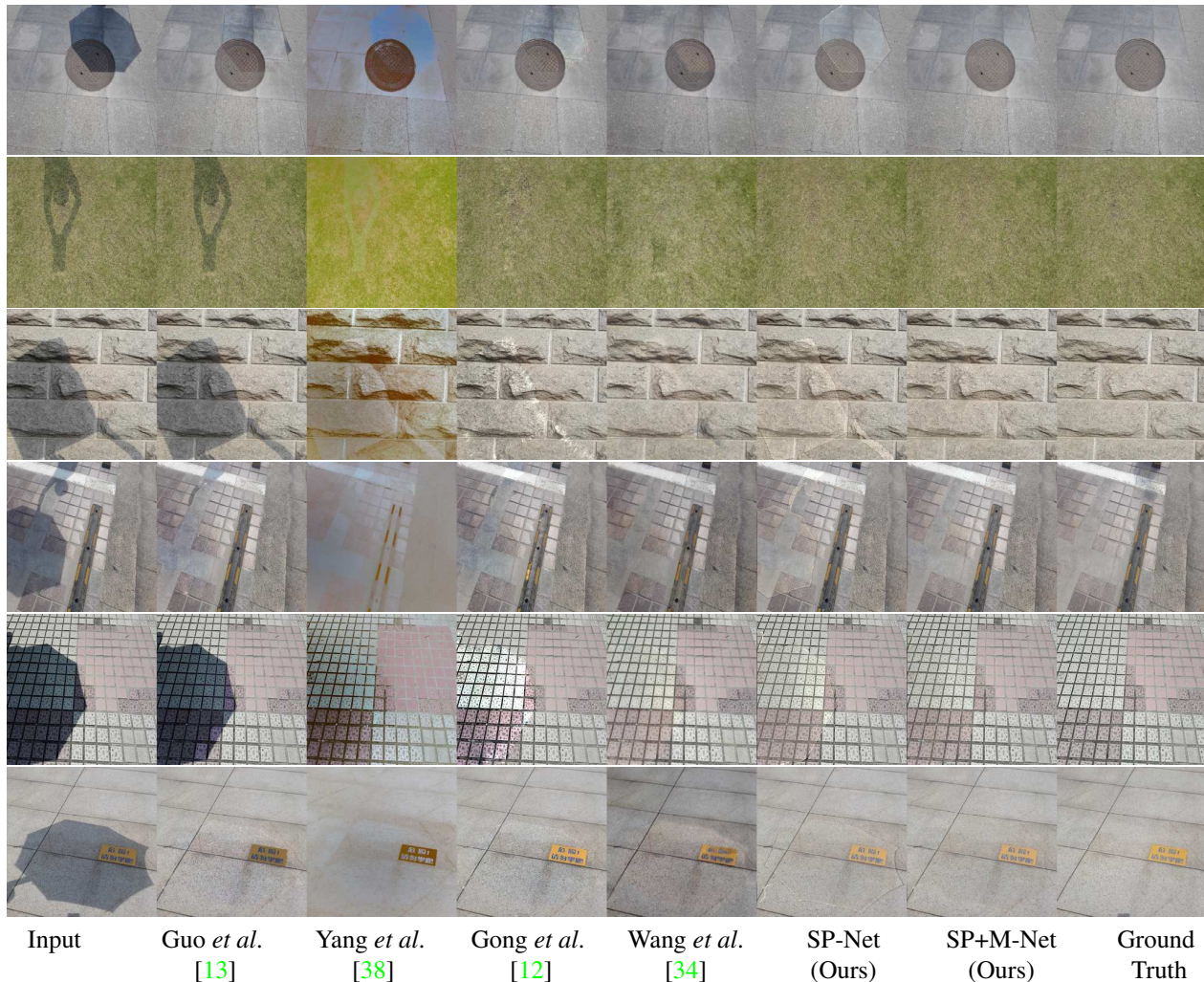


Figure 8: **Comparison of shadow removal on ISTD dataset.** Qualitative comparison between our method and previous state-of-the-art methods: Guo *et al.*[13], Yang *et al.*[38], Gong *et al.*[12], and Wang *et al.*[34]. “SP-Net” are the shadow removal results using the parameters computed from SP-Net and a binary shadow mask. “SP+M-Net” are the shadow removal results using the parameters computed from SP-Net and the shadow matte computed from M-Net.

6. Conclusions

In this work, we have presented a novel framework for shadow removal in single images. Our main contribution is to use deep networks as the parameters estimators for an illumination model. Our approach has advantages over previous approaches. Comparing to the traditional methods using an illumination model for removing shadows, our deep networks can estimate the parameters for the model from a single image accurately and automatically. Comparing to deep learning methods that perform shadow removal via an end-to-end mapping, our shadow removal framework outputs images with high quality and no artifact since we do not use the deep network to output the per-pixel values. Our model clearly achieves state-of-the-art shadow removal

results on the ISTD dataset. Our current approach can be extended in a number of ways. A more physically plausible illumination model would help the framework to output more realistic images. It would also be useful to develop a deep-learning based framework for shadow editing via a physical illumination model.

Acknowledgements. This work was partially supported by the NSF EarthCube program (Award 1740595), the National Neographic/Microsoft AI for Earth program, the Partner University Fund, the SUNY2020 Infrastructure Transportation Security Center, and a gift from Adobe. Computational support provided by the Institute for Advanced Computational Science and a GPU donation from NVIDIA. We thank Tomas Vicente for assistance with the manuscript.

References

- [1] Harry G. Barrow and J. Martin Tenenbaum. Recovering intrinsic scene characteristics from images. *Computer Vision Systems*, pages 3–26, 1978. [1](#), [2](#)
- [2] Joachim M. Buhmann. Weakly supervised structured output learning for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, CVPR '12, 2012. [7](#)
- [3] Yung-Yu Chuang, Dan B Goldman, Brian Curless, David H. Salesin, and Richard Szeliski. Shadow matting and compositing. *ACM Transactions on Graphics*, 22(3):494–500, July 2003. Special Issue of the SIGGRAPH 2003 Proceedings. [1](#), [2](#), [3](#)
- [4] R. Dennis Cook. Influential observations, high leverage points, and outliers in linear regression. *Statistical Science*, pages 393–397, 1986. [4](#)
- [5] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2009. [4](#)
- [6] Mark S. Drew. Recovery of chromaticity image free from shadows via illumination invariance. In *IEEE Workshop on Color and Photometric Methods in Computer Vision, ICCV03*, pages 32–39, 2003. [2](#), [4](#)
- [7] Graham Finlayson, Maryam Mohammadzadeh Darrodi, and Michal Mackiewicz. Rank-based camera spectral sensitivity estimation. *J. Opt. Soc. Am. A*, 33(4):589–599, Apr 2016. [3](#)
- [8] Graham Finlayson, Mark Drew, and Cheng Lu. Entropy minimization for shadow removal. *International Journal of Computer Vision*, 2009. [2](#), [4](#)
- [9] Graham Finlayson and Mark S. Drew. 4-sensor camera calibration for image representation invariant to shading, shadows, lighting, and specularities. In *Proceedings of the International Conference on Computer Vision*, volume 2, pages 473–480 vol.2, July 2001. [2](#)
- [10] Graham Finlayson, S.D. Hordley, Cheng Lu, and M.S. Drew. On the removal of shadows from images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006. [1](#), [2](#)
- [11] Graham Finlayson, Steven D. Hordley, and Mark S. Drew. Removing shadows from images. In *Proceedings of the European Conference on Computer Vision, ECCV '02*, pages 823–836, London, UK, UK, 2002. Springer-Verlag. [2](#), [4](#)
- [12] Han Gong and Darren Cosker. Interactive removal and ground truth for difficult shadow scenes. *J. Opt. Soc. Am. A*, 33(9):1798–1811, 2016. [2](#), [4](#), [6](#), [7](#), [8](#)
- [13] Ruiqi Guo, Qieyun Dai, and Derek Hoiem. Paired regions for shadow detection and removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012. [1](#), [2](#), [4](#), [6](#), [7](#), [8](#)
- [14] Xiang Huang, Gang Hua, J. Tumblin, and L. Williams. What characterizes a shadow boundary under the sun and sky? In *Proceedings of the International Conference on Computer Vision*, 2011. [1](#), [4](#)
- [15] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017. [2](#), [3](#), [6](#)
- [16] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of GANs for improved quality, stability, and variation. In *International Conference on Learning Representations*, 2018. [2](#)
- [17] Hieu Le, Bento Goncalves, Dimitris Samaras, and Heather Lynch. Weakly labeling the antarctic: The penguin colony case. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2019. [7](#)
- [18] Hieu Le, Tomas F. Yago Vicente, Vu Nguyen, Minh Hoai, and Dimitris Samaras. A+D Net: Training a shadow detector with adversarial shadow attenuation. In *Proceedings of the European Conference on Computer Vision*, 2018. [1](#), [7](#)
- [19] Hieu Le, Chen-Ping Yu, Gregory Zelinsky, and Dimitris Samaras. Co-localization with category-consistent features and geodesic distance propagation. In *ICCV 2017 Workshop on CEFRL: Compact and Efficient Feature Representation and Learning in Computer Vision*, 2017. [7](#)
- [20] A. Levin, D. Lischinski, and Y. Weiss. A closed-form solution to natural image matting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2):228–242, 2008. [2](#)
- [21] Si Liu, Jiashi Feng, Csaba Domokos, Hui Xu, Junshi Huang, Zhenzhen Hu, and Shuicheng Yan. Fashion parsing with weak color-category labels. *IEEE Transactions on Multimedia*, 16:253–265, 2014. [7](#)
- [22] Yang Liu, Zechao Li, Jinhui Tang, and Hanqing Lu. Weakly-supervised dual clustering for image semantic segmentation. *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2075–2082, 2013. [7](#)
- [23] Augustus Odena, Vincent Dumoulin, and Chris Olah. Deconvolution and checkerboard artifacts. *Distill*, 2016. [2](#)
- [24] Thomas Porter and Tom Duff. Compositing digital images. *Proceedings of the ACM SIGGRAPH Conference on Computer Graphics*, 18(3), January 1984. [2](#), [3](#)
- [25] Liangqiong Qu, Jiandong Tian, Shengfeng He, Yandong Tang, and Rynson W. H. Lau. Deshadownet: A multi-context embedding deep network for shadow removal. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017. [1](#), [3](#)
- [26] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention*, volume 9351 of *LNCS*, pages 234–241, 2015. [5](#)
- [27] Wang Shiting and Zheng Hong. Clustering-based shadow edge detection in a single color image. In *International Conference on Mechatronic Sciences, Electric Engineering and Computer*, pages 1038–1041, Dec 2013. [1](#)
- [28] Yael Shor and Dani Lischinski. The shadow meets the mask: Pyramid-based shadow removal. *Computer Graphics Forum*, 27(2):577–586, April 2008. [1](#), [2](#), [3](#), [4](#)
- [29] Ashish Shrivastava, Tomas Pfister, Oncel Tuzel, Josh Susskind, Wenda Wang, and Russ Webb. Learning from simulated and unsupervised images through adversarial training. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016. [7](#)
- [30] Alvy Ray Smith and James F. Blinn. Blue screen matting. In *Proceedings of the ACM SIGGRAPH Conference on Computer Graphics*, 1996. [3](#)

- [31] Tomas F. Yago Vicente, Minh Hoai, and Dimitris Samaras. Leave-one-out kernel optimization for shadow detection and removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(3):682–695, 2018. 2
- [32] Tomas F. Yago Vicente, Le Hou, Chen-Ping Yu, Minh Hoai, and Dimitris Samaras. Large-scale training of shadow detectors with noisily-annotated shadow examples. In *Proceedings of the European Conference on Computer Vision*, 2016. 1, 3, 6
- [33] Tomas F. Yago Vicente and Dimitris Samaras. Single image shadow removal via neighbor-based region relighting. In *Proceedings of the European Conference on Computer Vision Workshops*, 2014. 2
- [34] Jifeng Wang, Xiang Li, and Jian Yang. Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 1, 2, 3, 5, 6, 7, 8
- [35] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 2
- [36] Steve Wright. Digital compositing for film and video. In *Focal Press*, 2001. 3
- [37] Saining Xie, Ross Girshick, Piotr Dollar, Zhuowen Tu, and Kaiming He. Aggregated residual transformations for deep neural networks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017. 4
- [38] Qingxiong Yang, Kar Han Tan, and Narendra Ahuja. Shadow removal using bilateral filtering. *IEEE Transactions on Image Processing*, 21:4361–4368, 2012. 6, 7, 8
- [39] L. Zhang, Q. Zhang, and C. Xiao. Shadow remover: Image shadow removal based on illumination recovering optimization. *IEEE Transactions on Image Processing*, 24(11), Nov 2015. 1, 2
- [40] Richard Zhang, Phillip Isola, and Alexei A Efros. Colorful image colorization. In *ECCV*, 2016. 2, 6
- [41] Lei Zhu, Zijun Deng, Xiaowei Hu, Chi-Wing Fu, Xuemiao Xu, Jing Qin, and Pheng-Ann Heng. Bidirectional feature pyramid network with recurrent attention residual modules for shadow detection. In *Proceedings of the European Conference on Computer Vision*, 2018. 4, 6